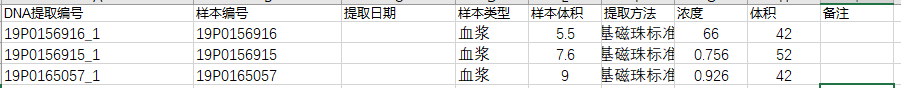
1. 数据疑问：

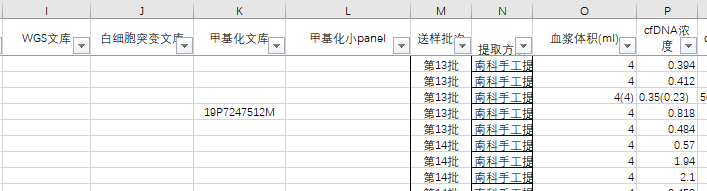
上海复旦中山胰腺癌 20191107.xlsx 文件

sheet1，没有提供诊断信息，但是在临床信息表中所有受试者的诊断都是‘胰腺癌’，OK？

原始数据中的医院编号 被重命名为 住院号；

批次效应：检测批次在标准数据库中没有，是否需要加入？





1. 根据项目组织并管理数据

不要在raw Data上做任何操作

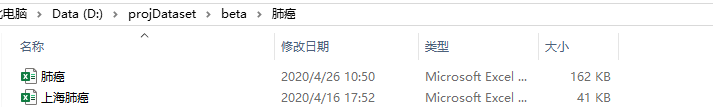
RawData整理需要有详细的说明

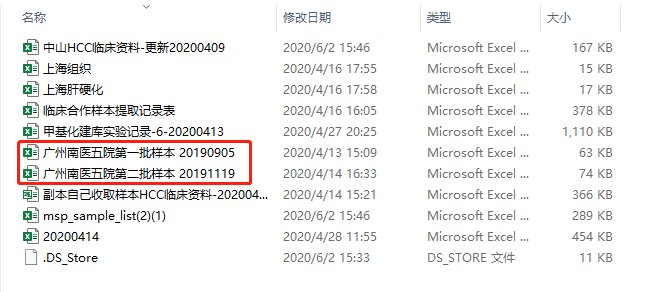
肝癌/肺癌/胃癌/肝癌/等，规范化rawdata文件的命名，中英文统一

比较准确的命名，有待优化：



不准确的命名





1. 文档的形式写明原始数据预处理到标准化数据库的修改过程。
2. 建议增加汇总的表格，对应样本与各文库之间的关系

下图来源于湖南直肠癌的临床数据：

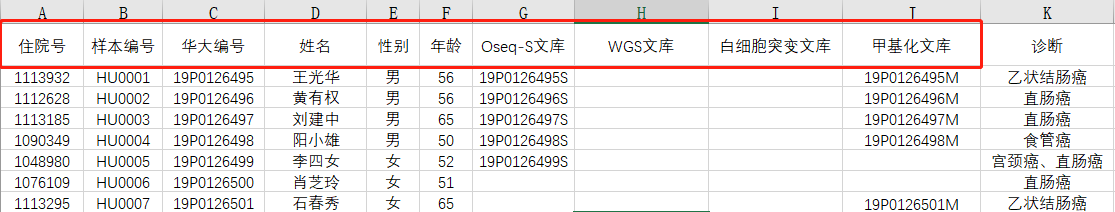
住院号：对应医疗记录；考虑到多中心，中心编号也需要加入住院号；形成每个受试者unique编号。

样本编号：对应生物样本

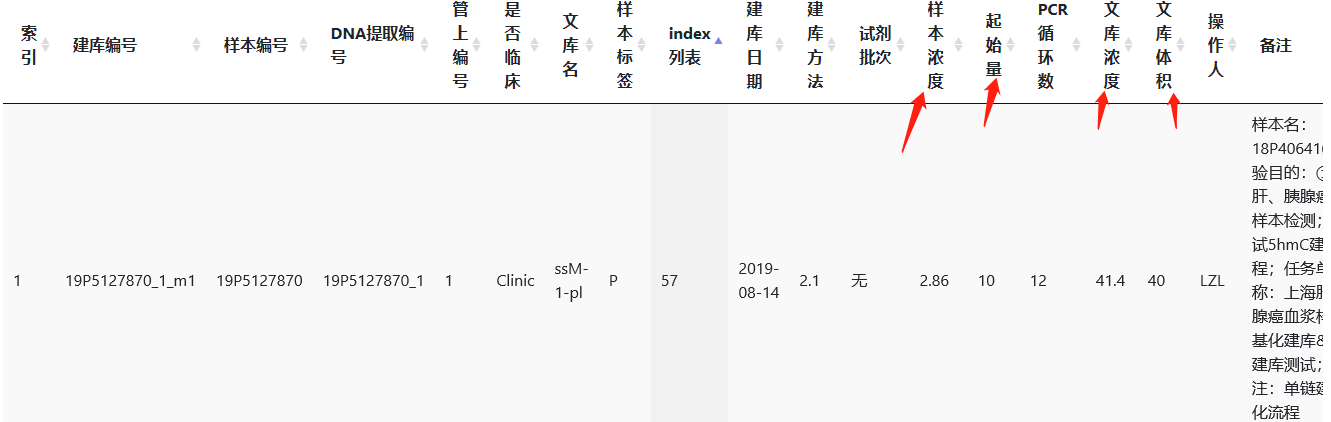
华大编号：对应？？

Oseq-S文库编号、WGS文库、白细胞突变文库和甲基化文库分别对应相应的试验结果

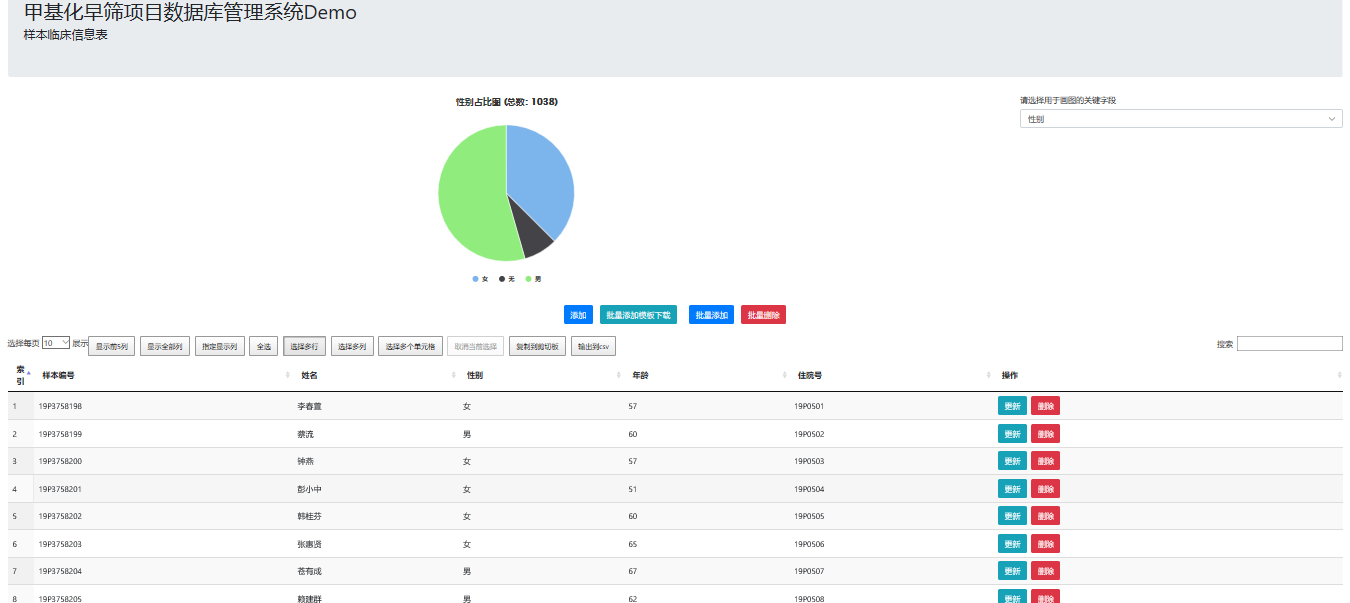
是否有其他的编号？



1. 部分字段没有相应的单位



1. 在系统中进行检索时，更关注的可能是查询出来的结果表，可视化的功能可以作为附加。



1. DNA样本提取信息和样本库存，可能是使用者的关注重点，DNA提取、使用记录和库存能不能组织成一个表，记录DNA信息；

样本临床信息表名称建议修改为 生物样本信息；

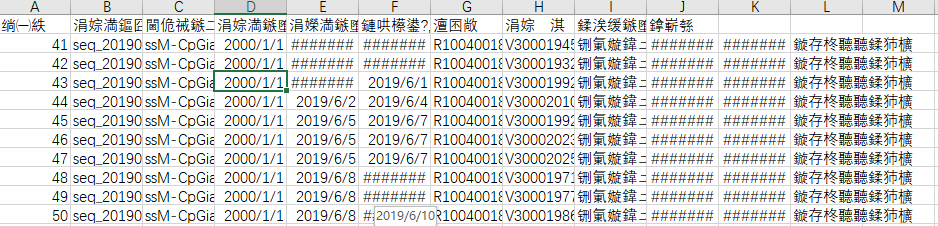


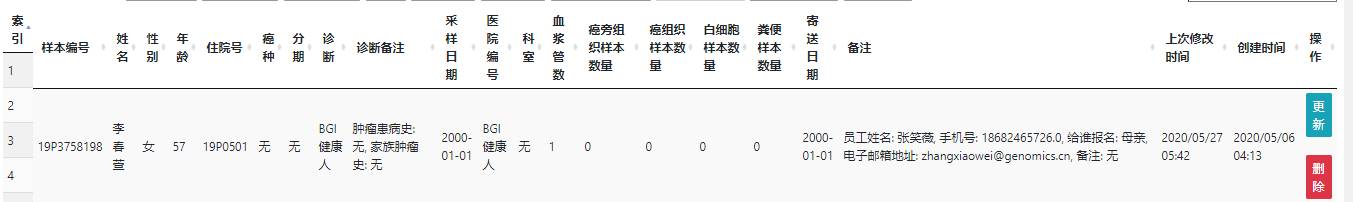
1. 目前的临床信息表，看起来更像是临床采集的生物样本信息；如果需要保留受试者的临床信息，可能需要更多加入：病史、肿瘤病灶记录、治疗史、实验室检查、手术记录、术后随访及治疗、病理记录、效果评估等；且随着后期数据和临床试验的增加，临床信息相关表格，具有可扩展性。

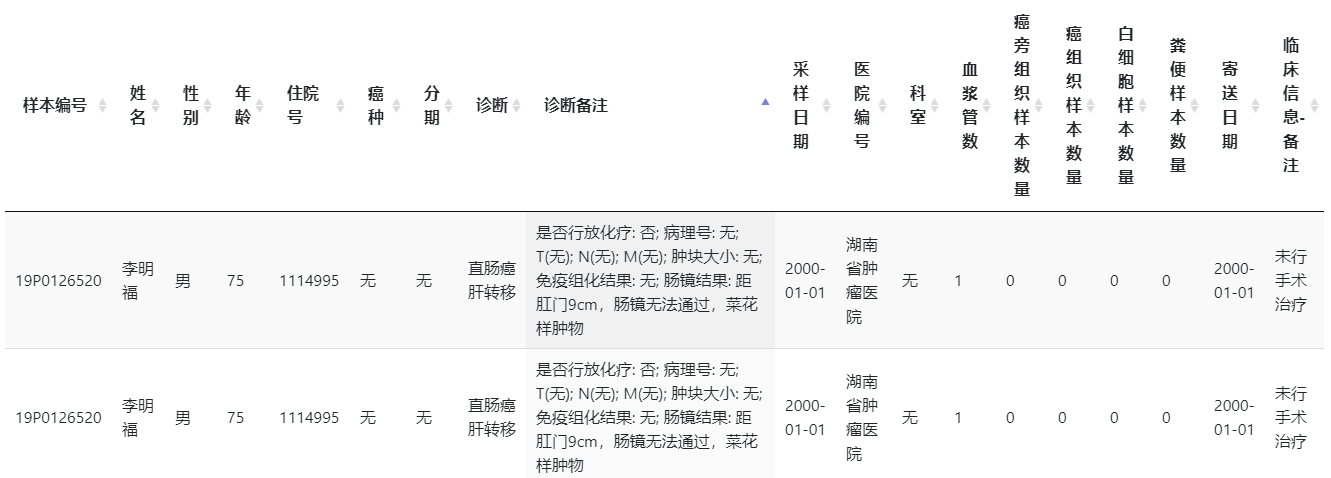
1.“备注”字段 内容在原始数据表中，联合体现为联系方式，且原始数据表中已有备注字段；建议这里字段 增设一个 “联系方式”，“备注”字段沿用原表内容。

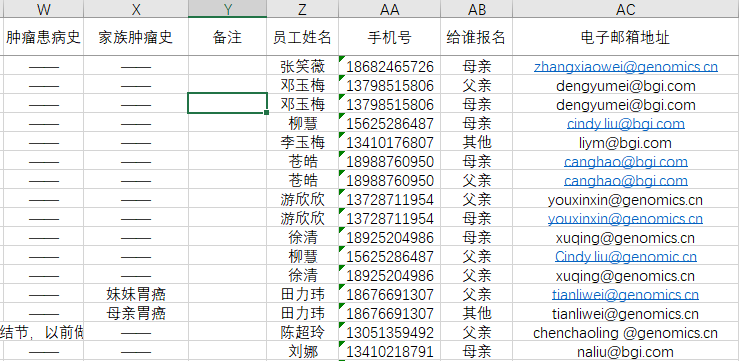
2.“诊断备注“字段 对健康人填写了病史信息/对肿瘤患者备注的病灶信息，建议另成表格记录；有重复记录。

3.输出到csv：乱码，只能输出10条记录？









1. 癌种名称标准化：放在哪个阶段？数据标准化入库or数据统计分析阶段；建议放在数统阶段，数据库中则提供最原始的记录。原因：纳入排除标准、后期人群分层研究。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 湖南直肠癌数据集 |  |  | 肝癌数据 |  |
| 癌肿名称 | 出现频次 |  | 癌肿名称 | 出现频次 |
| 腹部肿块查因 | 1 |  | 肝细胞肝癌 | 13 |
| 宫颈癌、直肠癌 | 1 |  | （肝）腺癌 | 1 |
| 横结肠息肉 | 1 |  | 差分化胆管癌 | 1 |
| 降结肠癌 | 1 |  | 差分化肝内胆管细胞癌 | 1 |
| 结肠癌 | 1 |  | 胆管内乳头状癌 | 1 |
| 升结肠癌 | 3 |  | 低分化腺癌 | 1 |
| 食管癌 | 1 |  | 肝内胆管癌 | 1 |
| 胃癌 | 1 |  | 肝内胆管细胞癌 | 8 |
| 乙状结肠癌 | 21 |  | 肝细胞癌 | 5 |
| 右半结肠癌 | 14 |  | 肝细胞肝癌 | 160 |
| 直肠癌 | 49 |  | 高分化肝细胞肝癌 | 2 |
| 直肠癌肝转移 | 1 |  | 合型肝细胞癌-胆管细胞癌 | 2 |
| 直肠癌术后 | 1 |  | 合性肝细胞癌和胆管癌 | 1 |
| 左半结肠癌 | 4 |  | 混合型肝细胞癌-胆管癌 | 1 |
|  |  |  | 混合型肝细胞癌-胆管细胞癌 | 3 |
|  |  |  | 混合性癌 | 1 |
|  |  |  | 混合性肝细胞肝癌和胆管细胞癌 | 1 |
|  |  |  | 间叶源性肿瘤 | 1 |
|  |  |  | 倾向差分化胆管癌 | 1 |
|  |  |  | 倾向低分化肝细胞肝癌 | 1 |
|  |  |  | 倾向肝细胞肝癌 | 1 |
|  |  |  | 倾向肝细胞腺瘤 | 1 |
|  |  |  | 上皮样细胞肿瘤 | 1 |
|  |  |  | 神经内分泌瘤 | 1 |
|  |  |  | 神经内分泌肿瘤 | 1 |
|  |  |  | 双表型肝细胞肝癌 | 1 |
|  |  |  | 无 | 1 |
|  |  |  | 腺癌 | 1 |
|  |  |  | 肿瘤性病变 | 1 |

